

# Attention Is All You Need

## Ein Natural-Language-Processing Lernpfad von Word-Embeddings bis zu den Transformern

**Einleitung:** Im letzten Jahrzehnt haben auf dem Gebiet der Verarbeitung natürlicher Sprache (Natural Language Processing, NLP) gleich zwei Revolutionen stattgefunden. Das aufwändige Feature Engineering der Sprache durch Menschen wurde von Computern mit Deep Learning übernommen. Es wurden neue Algorithmen entwickelt, mit denen ein Computer die Bedeutung von Sprache erfassen und benutzen kann. Komplexe neuronale Netzwerke sind in der Lage, die Zusammenhänge von menschlicher Sprache zu modellieren. Nur wenige Jahre später haben die Transformer, eine neue Deep Learning Architektur, die bisherigen Modelle praktisch verdrängt. Die Transformer ermöglichen die sehr guten Übersetzungen, die Programme wie Google Translate oder DeepL liefern. Diese Arbeit stellt die verschiedenen Technologien der letzten Jahre in Form eines Lernpfades in Jupyter Notebooks vor, die NLP verändert haben.

**Aufgabenstellung:** Es sind maschinelle Übersetzer in verschiedenen Konfigurationen zu trainieren, um die Effekte von 'Attention' und 'vortrainierten Embeddings' zu demonstrieren. Beim Übersetzen sind nicht alle Wörter gleich wichtig, bzw. an unterschiedlichen Stellen sind unterschiedliche Wörter des zu übersetzenden Satzes wichtig. Der 'Attention-Mechanismus' erlaubt dem Netzwerk, sich auf die wichtigen Wörter zu konzentrieren. Der Computer kann eine numerische Repräsentation, Embeddings, eines Wortes lernen. Diese Embeddings können während des Trainierens des Netzwerkes gelernt werden oder durch bereits trainierte Embeddings (z. B. von Google, Facebook) ersetzt werden.

**Ergebnis:** Die Übersetzung mit Attention ist bedeutend besser als ohne Attention. Die Semantik

innerhalb eines Satzes wird besser erfasst und Details werden viel besser in die Zielsprache übersetzt. Die vortrainierten Embeddings haben die Übersetzung verschlechtert. Das Modell mit Attention und der Transformer liefern bei eher einfachen Sätzen meist gute Übersetzungen.

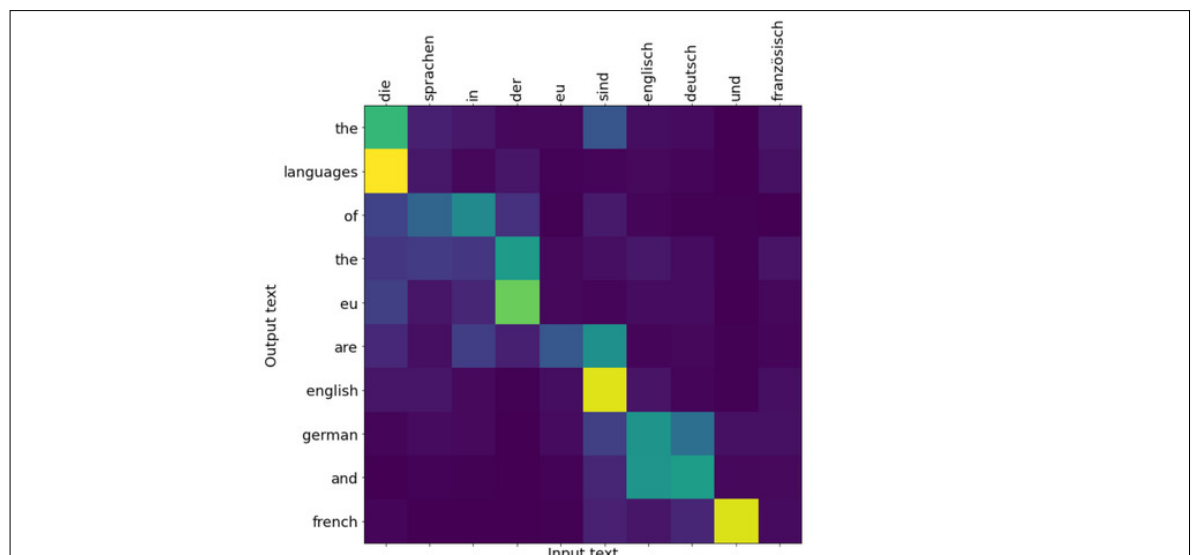
**Beispiel: Ein Satz wird von allen trainierten Modellen übersetzt**  
Eigene Darstellung

Modell	Übersetzung
Deutsch	Ab 1985 wurden mit dem Schengener Übereinkommen die Binnengrenzen zwischen den Mitgliedsländern geöffnet.
Ohne Attention	from 1985 the schengen convention opened up the possibility of internal border controls between member states
Attention	in 1985 the schengen agreement opened up internal borders between the member states
Attention, Embeddings	the schengen agreement concluded with the schengen agreement between the member states
Transformer	from 1985 , the schengen agreement opened the internal borders between member states .

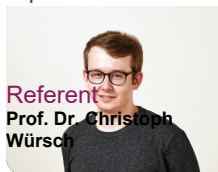
**Beispiel: Ein Satz wird von allen trainierten Modellen übersetzt**  
Eigene Darstellung

Modell	Übersetzung
Deutsch	Die verbreitetsten Sprachen in der EU sind Englisch, Deutsch und Französisch.
Ohne Attention	the english language is in english english and french
Attention	the languages of the eu are english german and french
Attention, Embeddings	the language of the eu is book book book and french literature
Transformer	the most widespread languages in the eu are english , german and french .

**Mit Attention kann ausgegeben werden, auf welche Wörter sich das Netzwerk konzentriert, wenn ein Wort übersetzt wird**  
Eigene Darstellung



Diplomand



Referent  
Prof. Dr. Christoph Würsch

Ko-Referent  
Dr. Shao Jü Woo

Themengebiet  
Computational Engineering, Ingenieurinformatik